AD NO. ____________________________
DTC PROJECT NO. 8-CO-160-UXO-021
REPORT NO. ATC-9089

STANDARDIZED

UXO TECHNOLOGY DEMONSTRATION SITE

OPEN FIELD SCORING RECORD NO. 673

SITE LOCATION:
U.S. ARMY ABERDEEN PROVING GROUND

DEMONSTRATOR:
NAVAL RESEARCH LABORATORIES (NRL)
CODE 6110 NAVAL RESEARCH LABORATORIES
WASHINGTON, DC 20375-5342

TECHNOLOGY TYPE/PLATFORM:
MAGNETOMETER MTADS/TOWED

PREPARED BY:
U.S. ARMY ABERDEEN TEST CENTER
ABERDEEN PROVING GROUND, MD 21005-5059

SEPTEMBER 2005

Prepared for:
U.S. ARMY ENVIRONMENTAL CENTER
ABERDEEN PROVING GROUND, MD 21010-5401

U.S. ARMY DEVELOPMENTAL TEST COMMAND
ABERDEEN PROVING GROUND, MD 21005-5055

DISTRIBUTION UNLIMITED, SEPTEMBER 2005.

DISPOSITION INSTRUCTIONS

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
September 2005

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
21 and 22 June 2004

4. TITLE AND SUBTITLE
STANDARDIZED UXO TECHNOLOGY DEMONSTRATION SITE OPEN FIELD SCORING RECORD NO. 673 (NAVAL RESEARCH LABORATORIES)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Overbay, Larry; Robitaille, George
The Standardized UXO Technology Demonstration Site Scoring Committee

5d. PROJECT NUMBER
8-CO-160-UXO-021

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Aberdeen Test Center
ATTN: CSTE-DTC-AT-SL-E
Aberdeen Proving Ground, MD 21005-5059

8. PERFORMING ORGANIZATION REPORT NUMBER
ATC-9089

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Environmental Center
ATTN: SFIM-AEC-ATT
Aberdeen Proving Ground, MD 21005-5401

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Same as Item 8

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This scoring record documents the efforts of Naval Research Laboratories to detect and discriminate inert unexploded ordnance (UXO) utilizing the APG Standardized UXO Technology Demonstration Site Open Field. Scoring Records have been coordinated by Larry Overbay and the Standardized UXO Technology Demonstration Site Scoring Committee. Organizations on the committee include, the U.S. Army Corps of Engineers, the Environmental Security Technology Certification Program, the Strategic Environmental Research and Development Program, the Institute for Defense Analysis, the U.S. Army Environmental Center, and the U.S. Army Aberdeen Test Center.

15. SUBJECT TERMS
Naval Research Laboratories, UXO Standardized Technology Demonstration Site Program, Open Field, Magnetometer MTADS/towed

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UL		19b. TELEPHONE NUMBER *(Include area code)*

ACKNOWLEDGEMENTS

Authors:

Larry Overbay Jr.
Matthew Boutin
Military Environmental Technology Demonstration Center (METDC)
U.S. Army Aberdeen Test Center (ATC)
U.S. Army Aberdeen Proving Ground (APG)

Rick Fling
Aberdeen Test and Support Services (ATSS)
Sverdrup Technology, Inc.
U.S. Army Aberdeen Proving Ground (APG)

Christina McClung
Aberdeen Data Services Team (ADST)
Tri-S, Inc.
U.S. Army Aberdeen Proving Ground (APG)

Contributor:

George Robitaille
U.S. Army Environmental Center (AEC)
U.S. Army Aberdeen Proving Ground (APG)

TABLE OF CONTENTS

PAGE

SECTION 4. TECHNICAL PERFORMANCE RESULTS

PAGE

SECTION 5. ON-SITE LABOR COSTS

SECTION 6. COMPARISON OF RESULTS TO BLIND GRID DEMONSTRATION

SECTION 7. APPENDIXES

SECTION 1. GENERAL INFORMATION

1.1 BACKGROUND

Technologies under development for the detection and discrimination of unexploded ordnance (UXO) require testing so that their performance can be characterized. To that end, Standardized Test Sites have been developed at Aberdeen Proving Ground (APG), Maryland and U.S. Army Yuma Proving Ground (YPG), Arizona. These test sites provide a diversity of geology, climate, terrain, and weather as well as diversity in ordnance and clutter. Testing at these sites is independently administered and analyzed by the government for the purposes of characterizing technologies, tracking performance with system development, comparing performance of different systems, and comparing performance in different environments.

The Standardized UXO Technology Demonstration Site Program is a multi-agency program spearheaded by the U.S. Army Environmental Center (AEC). The U.S. Army Aberdeen Test Center (ATC) and the U.S. Army Corps of Engineers Engineering Research and Development Center (ERDC) provide programmatic support. The program is being funded and supported by the Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP) and the Army Environmental Quality Technology Program (EQT).

1.2 SCORING OBJECTIVES

The objective in the Standardized UXO Technology Demonstration Site Program is to evaluate the detection and discrimination capabilities of a given technology under various field and soil conditions. Inert munitions and clutter items are positioned in various orientations and depths in the ground.

The evaluation objectives are as follows:

a. To determine detection and discrimination effectiveness under realistic scenarios that vary targets, geology, clutter, topography, and vegetation.

b. To determine cost, time, and manpower requirements to operate the technology.

c. To determine demonstrator's ability to analyze survey data in a timely manner and provide prioritized "Target Lists" with associated confidence levels.

d. To provide independent site management to enable the collection of high quality, ground-truth, geo-referenced data for post-demonstration analysis.

1.2.1 Scoring Methodology

a. The scoring of the demonstrator's performance is conducted in two stages. These two stages are termed the RESPONSE STAGE and DISCRIMINATION STAGE. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver-operating

characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of false positive (P_{fp}), and those that do not correspond to any known item, termed background alarms.

b. The RESPONSE STAGE scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate ordnance from other anomalies. For the blind grid RESPONSE STAGE, the demonstrator provides the scoring committee with a target response from each and every grid square along with a noise level below which target responses are deemed insufficient to warrant further investigation. This list is generated with minimal processing and, since a value is provided for every grid square, will include signals both above and below the system noise level.

c. The DISCRIMINATION STAGE evaluates the demonstrator's ability to correctly identify ordnance as such and to reject clutter. For the blind grid DISCRIMINATION STAGE, the demonstrator provides the scoring committee with the output of the algorithms applied in the discrimination-stage processing for each grid square. The values in this list are prioritized based on the demonstrator's determination that a grid square is likely to contain ordnance. Thus, higher output values are indicative of higher confidence that an ordnance item is present at the specified location. For digital signal processing, priority ranking is based on algorithm output. For other discrimination approaches, priority ranking is based on human (subjective) judgment. The demonstrator also specifies the threshold in the prioritized ranking that provides optimum performance, (i.e. that is expected to retain all detected ordnance and rejects the maximum amount of clutter).

d. The demonstrator is also scored on EFFICIENCY and REJECTION RATIO, which measures the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of ordnance detections from the anomaly list, while rejecting the maximum number of anomalies arising from non-ordnance items. EFFICIENCY measures the fraction of detected ordnance retained after discrimination, while the REJECTION RATIO measures the fraction of false alarms rejected. Both measures are defined relative to performance at the demonstrator-supplied level below which all responses are considered noise, i.e., the maximum ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate.

e. Based on configuration of the ground truth at the standardized sites and the defined scoring methodology, there exists the possibility of having anomalies within overlapping halos and/or multiple anomalies within halos. In these cases, the following scoring logic is implemented:

(1) In situations where multiple anomalies exist within a single R_{halo}, the anomaly with the strongest response or highest ranking will be assigned to that particular ground truth item.

(2) For overlapping R_{halo} situations, ordnance has precedence over clutter. The anomaly with the strongest response or highest ranking that is closest to the center of a particular ground truth item gets assigned to that item. Remaining anomalies are retained until all matching is complete.

(3) Anomalies located within any R_{halo} that do not get associated with a particular ground truth item are thrown out and are not considered in the analysis.

f. All scoring factors are generated utilizing the Standardized UXO Probability and Plot Program, version 3.1.1.

1.2.2 Scoring Factors

Factors to be measured and evaluated as part of this demonstration include:

a. Response Stage ROC curves:

(1) Probability of Detection (P_d^{res}).

(2) Probability of False Positive (P_{fp}^{res}).

(3) Background Alarm Rate (BAR^{res}) or Probability of Background Alarm (P_{BA}^{res}).

b. Discrimination Stage ROC curves:

(1) Probability of Detection (P_d^{disc}).

(2) Probability of False Positive (P_{fp}^{disc}).

(3) Background Alarm Rate (BAR^{disc}) or Probability of Background Alarm (P_{BA}^{disc}).

c. Metrics:

(1) Efficiency (E).

(2) False Positive Rejection Rate (R_{fp}).

(3) Background Alarm Rejection Rate (R_{BA}).

d. Other:

(1) Probability of Detection by Size and Depth.

(2) Classification by type (i.e., 20-, 40-, 105-mm, etc.).

(3) Location accuracy.

(4) Equipment setup, calibration time and corresponding man-hour requirements.

(5) Survey time and corresponding man-hour requirements.

(6) Reacquisition/resurvey time and man-hour requirements (if any).

(7) Downtime due to system malfunctions and maintenance requirements.

1.3 STANDARD AND NONSTANDARD INERT ORDNANCE TARGETS

The standard and nonstandard ordnance items emplaced in the test areas are listed in Table 1. Standardized targets are members of a set of specific ordnance items that have identical properties to all other items in the set (caliber, configuration, size, weight, aspect ratio, material, filler, magnetic remanence, and nomenclature). Nonstandard targets are inert ordnance items having properties that differ from those in the set of standardized targets.

TABLE 1. INERT ORDNANCE TARGETS

Standard Type	Nonstandard (NS)
20-mm Projectile M55	20-mm Projectile M55
	20-mm Projectile M97
40-mm Grenades M385	40-mm Grenades M385
40-mm Projectile MKII Bodies	40-mm Projectile M813
BDU-28 Submunition	
BLU-26 Submunition	
M42 Submunition	
57-mm Projectile APC M86	
60-mm Mortar M49A3	60-mm Mortar (JPG)
	60-mm Mortar M49
2.75-inch Rocket M230	2.75-inch Rocket M230
	2.75-inch Rocket XM229
MK 118 ROCKEYE	
81-mm Mortar M374	81-mm Mortar (JPG)
	81-mm Mortar M374
105-mm Heat Rounds M456	
105-mm Projectile M60	105-mm Projectile M60
155-mm Projectile M483A1	155-mm Projectile M483A
	500-lb Bomb

JPG = Jefferson Proving Ground
HEAT = high-explosive, antitank

SECTION 2. DEMONSTRATION

2.1 DEMONSTRATOR INFORMATION

2.1.1 Demonstrator Point of Contact (POC) and Address

POC: Herb Nelson
202-767-3686
herb.nelson@nrl.navy.mil

Address: Naval Research Laboratory
Code 6110
Naval Research Laboratory
Washington, DC 20375-5342

2.1.2 System Description (provided by demonstrator)

The Multi-Sensor Towed Array Detection Systems (MTADS) hardware consists of a low-magnetic signature vehicle that is used to tow linear arrays of magnetometer and pulsed-induction sensors to conduct surveys of large areas to detect buried UXO (fig. 1). The MTADS tow vehicle, manufactured by Chenowth Racing Vehicles, is a custom-built off-road vehicle, specifically modified to have an extremely low magnetic signature. Most ferrous components have been removed from the body, drive train, and engine and replaced with nonferrous alloys.

Figure 1. Demonstrator system, MAG MTADS/towed.

The MTADS magnetometers are Cs-vapor full-field magnetometers (Geometrics Model 822ROV). Eight 1.75 meter long sensors are deployed as a linear magnetometer array.

The pulsed-induction sensors (specially modified Geonics EM61s for the baseline system and EM61 MKII's for this demonstration) are deployed as an overlapping array of three sensors. The sensors employed by MTADS have been modified to make them more compatible with vehicular speeds and to increase their sensitivity to small objects. The MTADS baseline EM61s have the sample gate at the earliest possible time. This enhances signal levels, and thus detection performance, but at the cost of classification ability. The EM61 MKIIs that will be evaluated were equipped with four sample gates. This is intended to enhance our ability to discriminate large objects from a collection of smaller fragments.

The sensor positions are measured in real-time (5 Hz) using the latest real time kinematic (RTK) Global Positioning System (GPS) technology. All navigation and sensor data are time-stamped and recorded by the data acquisition computer in the tow vehicle. The Data Analysis System (DAS) employs routines to convert these sensor and position data streams into anomaly maps for analysis.

2.1.3 Data Processing Description (provided by demonstrator)

The MTADS magnetometer array is pulled by the MTADS tow vehicle over the site at approximately 6 miles per hour. Lane spacing is the width of the MTADS tow vehicle, approximately 1.75 meters. Data are recorded from the array at 50 Hz. This results in a down-track sampling interval of $\approx$6 cm and a cross-track sampling interval of 25 cm. The EM61 sensors are arranged in an overlapping configuration as shown in Figure 1. Nominal survey speed is 3 mph and the sensor readings are recorded at 10 Hz. This results in a down-track sampling of $\approx$15 cm and a cross-track interval of 50 cm. In order to obtain sufficient views of the targets, we collect data in two orthogonal surveys.

Individual sensors in the electromagnetic (EM) array are located using a three-receiver RTK GPS system. From this set of receivers, we record the position of the master antenna at 20 Hz, and the vectors to the other two antennae at 10 Hz. All positions are recorded at full RTK precision, $\approx$ 2-5 cm. Since the magnetometer sensors are arranged in a rigid array with the GPS antenna hard mounted on the array, a single GPS measurement suffices. All sensor readings are referenced to the GPS 1-pulse per second (PPS) output so we are able to take full advantage of the precision of the GPS measurements.

The individual data streams (sensor readings, GPS positions, times, etc.) are collected by the data acquisition computer, running a custom variant of the MagLog NT program, and are each recorded in a separate file. These individual data files, which share a root name, include two (magnetometer array) or three (EM array) sensor data files and two (magnetometer array) or four (EM array) GPS files (one containing the NMEA GGK sentences corresponding to the position of the master antenna and an automatic voltage regulator (AVR) sentence giving one of the vectors to the secondary antennas, another containing the second AVR sentence, a third containing the UTC time tag, and the fourth containing the computer time-stamped arrival of the GPS PPS). All files are American Standard Code for Information Interchange (ASCII) format.

All files are transferred to the DAS using ZIP-250 disks. They are then checked for data quality, leveled, and the position information is applied to the sensor files. The result is a sequence of positioned measurements of the measured response. The latter file is what we refer to as raw data.

2.1.4 Data Submission Format

Data were submitted for scoring in accordance with data submission protocols outlined in the Standardized UXO Technology Demonstration Site Handbook. These submitted data are not included in this report in order to protect ground truth information.

2.1.5 Demonstrator Quality Assurance (QA) and Quality Control (QC) (provided by demonstrator)

There are two items that need to be checked daily to ensure adequate system performance: individual sensor response and reliability of GPS positions. Before beginning survey work each day, the performance of each of the sensors in the array is measured (after a 10 to 15 minute warm-up) by presenting a standard target to each sensor in turn. The resulting signals are checked against standard values.

Our data acquisition system gives the vehicle operator a continuous reading of the quality of the GPS fix. Our standard procedure is to take only data with a GPS fix quality of three (RTK fixed) or Figure 1 - Sketch of the MTADS EM61 array 2 (RTK float) and a precision dilution of precision (PDOP) of four or less. Before arriving at the site each day, we use standard GPS planning software to calculate the number of satellites that will be visible to the receivers and the PDOP achievable minute-by-minute throughout the day. This allows us to plan on short breaks during periods of poor satellite availability and keeps us from inadvertently taking data that will have to be discarded later. Another important feature of this GPS planning is the ability to take into account areas of restricted sky view (such as along the tree line at one edge of the APG site). In our experience, there is usually a brief period each day, on the order of 20 to 30 minutes, when good fixes can be obtained in even the most difficult environments. With planning, the system can be poised by the tree line ready to take data when the appropriate satellite alignment occurs.

At the end of each one-hour survey session, all survey data is transferred to the field data analyst for preliminary data quality checks. This process involves plotting the actual survey path as logged in the GPS files (color-coded by GPS fix quality) to ensure that GPS data of sufficient quality was obtained during the survey. Following this, the individual sensor files are examined for completeness and consistency. It is at this stage that any sensor malfunctions, drifts, etc. are flagged and reported to the field crew for correction. The final task for the field analyst is to calculate a position for each sensor reading and apply it to the reading. The mapped data files are then ready for analysis either in the field, or at a later time.

2.1.6 Additional Records

The following record(s) by this vendor can be accessed via the Internet as MicroSoft Word documents at www.uxotestsites.org. The Blind Grid counterpart to this report is Scoring Record #671.

2.2 APG SITE INFORMATION

2.2.1 Location

The APG Standardized Test Site is located within a secured range area of the Aberdeen Area. The Aberdeen Area of APG is located approximately 30 miles northeast of Baltimore at the northern end of the Chesapeake Bay. The Standardized Test Site encompasses 17 acres of upland and lowland flats, woods and wetlands.

2.2.2 Soil Type

According to the soils survey conducted for the entire area of APG in 1998, the test site consists primarily of Elkton Series type soil (ref 2). The Elkton Series consist of very deep, slowly permeable, poorly drained soils. These soils formed in silty aeolin sediments and the underlying loamy alluvial and marine sediments. They are on upland and lowland flats and in depressions of the Mid-Atlantic Coastal Plain. Slopes range from 0 to 2 percent.

ERDC conducted a site-specific analysis in May of 2002 (ref 3). The results basically matched the soil survey mentioned above. Seventy percent of the samples taken were classified as silty loam. The majority (77 percent) of the soil samples had a measured water content between 15- and 30-percent with the water content decreasing slightly with depth.

For more details concerning the soil properties at the APG test site, go to www.uxotestsites.org on the web to view the entire soils description report.

2.2.3 Test Areas

A description of the test site areas at APG is included in Table 2.

TABLE 2. TEST SITE AREAS

Area	Description
Calibration Grid	Contains 14 standard ordnance items buried in six positions at various angles and depths to allow demonstrator to calibrate their equipment.
Blind Test Grid	Contains 400 grid cells in a 0.2-hectare (0.5 acre) site. The center of each grid cell contains ordnance, clutter or nothing.
Open Field	A 4-hectare (10-acre) site containing open areas, dips, ruts and obstructions that challenge platform systems or hand held detectors. The challenges include a gravel road, wet areas and trees. The vegetation height varies from 15 to 25 cm.

SECTION 3. FIELD DATA

3.1 DATE OF FIELD ACTIVITIES (21 and 22 June 2004)

3.2 AREAS TESTED/NUMBER OF HOURS

Areas tested and total number of hours operated at each site are summarized in Table 3.

TABLE 3. AREAS TESTED AND NUMBER OF HOURS

Area	Number of Hours
Calibration Lanes	0.75
Open Field	7.23

3.3 TEST CONDITIONS

3.3.1 Weather Conditions

An APG weather station located approximately one mile west of the test site was used to record average temperature and precipitation on a half hour basis for each day of operation. The temperatures listed in Table 4 represent the average temperature during field operations from 0700 to 1700 hours while precipitation data represents a daily total amount of rainfall. Hourly weather logs used to generate this summary are provided in Appendix B.

TABLE 4. TEMPERATURE/PRECIPITATION DATA SUMMARY

Date, 2004	Average Temperature, °F	Total Daily Precipitation, in.
21 June	74.14	0.00
22 June	79.78	0.24

3.3.2 Field Conditions

NRL surveyed the Open Field 21 and 22 June 2004. The field was wet and had areas of standing water throughout the Open Field.

3.3.3 Soil Moisture

Three soil probes were placed at various locations within the site to capture soil moisture data: Blind Grid, Calibration, Mogul, and Wooded areas. Measurements were collected in percent moisture and were taken twice daily (morning and afternoon) from five different soil depths (1 to 6 in., 6 to 12 in., 12 to 24 in., 24 to 36 in., and 36 to 48 in.) from each probe. Soil moisture logs are included in Appendix C.

3.4 FIELD ACTIVITIES

3.4.1 Setup/Mobilization

These activities included initial mobilization and daily equipment preparation and break down. A three-person crew took 1 hour and 30 minutes to perform the initial setup and mobilization. There was 20 minutes of daily equipment preparation and end of the day equipment break down lasted 25 minutes.

3.4.2 Calibration

NRL spent a total of 45 minutes in the calibration lanes, of which 27 minutes was spent collecting data.

3.4.3 Downtime Occasions

Occasions of downtime are grouped into five categories: equipment/data checks or equipment maintenance, equipment failure and repair, weather, Demonstration Site issues, or breaks/lunch. All downtime is included for the purposes of calculating labor costs (section 5) except for downtime due to Demonstration Site issues. Demonstration Site issues, while noted in the Daily Log, are considered non-chargeable downtime for the purposes of calculating labor costs and are not discussed. Breaks and lunches are discussed in this section and billed to the total Site Survey area.

3.4.3.1 Equipment/data checks, maintenance. Equipment data checks and maintenance activities accounted for 28 minutes of site usage time. These activities included changing out batteries and routine data checks to ensure the data was being properly recorded/collected. NRL spent an additional 57 minutes for breaks and lunches.

3.4.3.2 Equipment failure or repair. No time was needed to resolve equipment failures that occurred while surveying the Open Field.

3.4.3.3 Weather. No weather delays occurred during the survey.

3.4.4 Data Collection

NRL spent a total time of 7 hours and 14 minutes in the Open Field area, 5 hours and 4 minutes of which was spent collecting data.

3.4.5 Demobilization

The NRL survey crew conducted a full demonstration of the Open Field site. Therefore, demobilization did not occur until 22 June 2004. On that day, it took the crew 1 hour to break down and pack up their equipment.

3.5 PROCESSING TIME

NRL submitted the raw data from the demonstration activities on the last day of the demonstration, as required. The scoring submittal data was also provided within the required 30-day timeframe.

3.6 DEMONSTRATOR'S FIELD PERSONNEL

Herb Nelson, NRL
Dan Steinhurst, NOVA Research, Inc.
Glenn Harbaugh, NOVA Research, Inc.

3.7 DEMONSTRATOR'S FIELD SURVEYING METHOD

NRL surveyed long lines from one side of the sight to the other starting at the Southwest corner of the Open Field area going south to north.

3.8 SUMMARY OF DAILY LOGS

Daily logs capture all field activities during this demonstration and are located in Appendix D. Activities pertinent to this specific demonstration are indicated in highlighted text.

SECTION 4. TECHNICAL PERFORMANCE RESULTS

4.1 ROC CURVES USING ALL ORDNANCE CATEGORIES

Figure 2 shows the probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of false positive. Figure 3 shows both probabilities plotted against their respective background alarm rate. Both figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination. Note that all points have been rounded to protect the ground truth.

The overall ground truth is composed of ferrous and non-ferrous anomalies. Due to limitations of the magnetometer, the non-ferrous items cannot be detected. Therefore, the ROC curves presented in this section are based on the subset of the ground truth that is solely made up of ferrous anomalies.

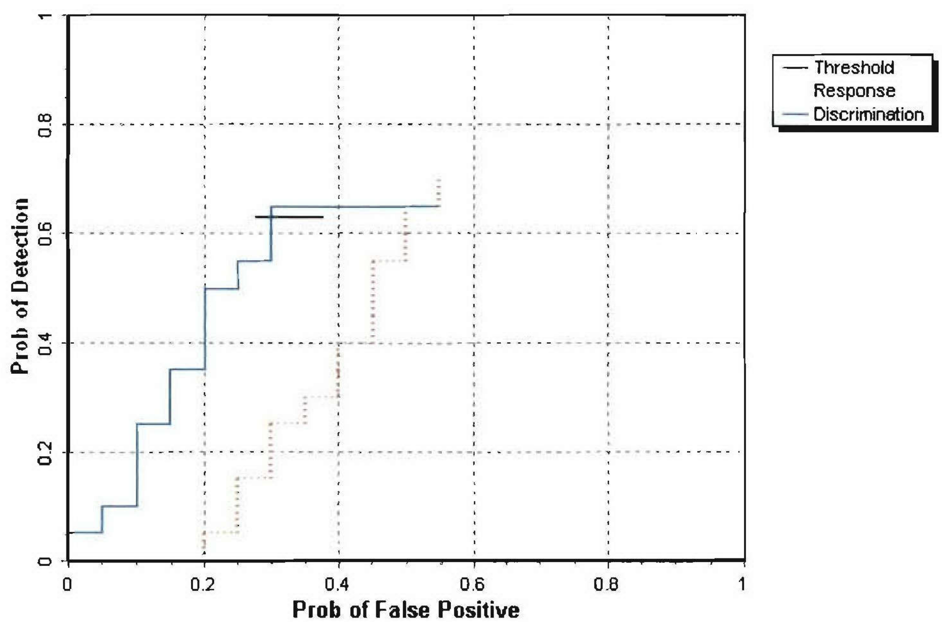

Figure 2. MTADS MAG/towed open field probability of detection for response and discrimination stages versus their respective probability of false positive over all ordnance categories combined.

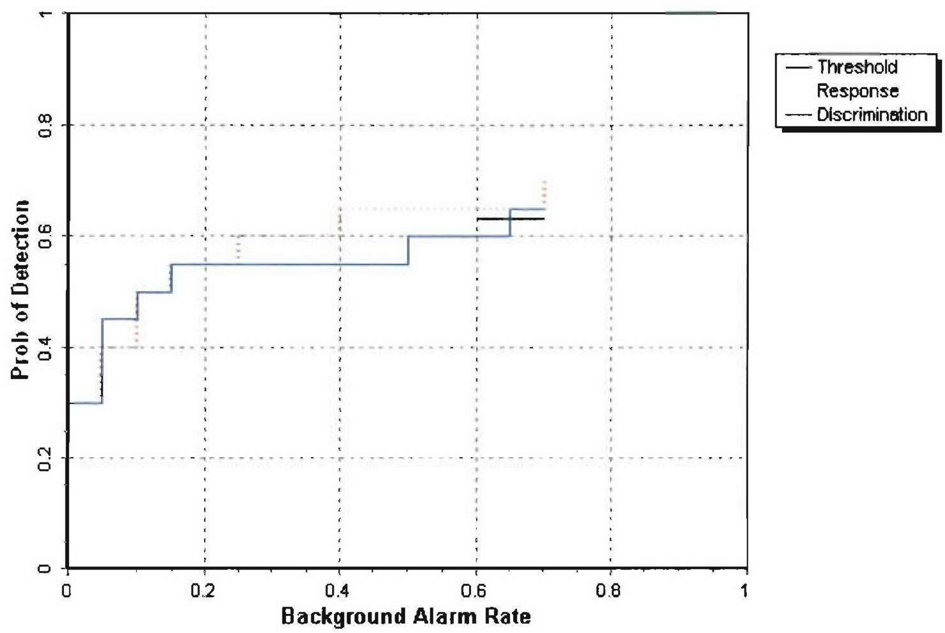

Figure 3. MTADS MAG/towed open field probability of detection for response and discrimination stages versus their respective background alarm rate over all ordnance categories combined.

4.2 ROC CURVES USING ORDNANCE LARGER THAN 20 MM

Figure 4 shows the probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of false positive when only targets larger than 20 mm are scored. Figure 5 shows both probabilities plotted against their respective background alarm rate. Both figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination. Note that all points have been rounded to protect the ground truth.

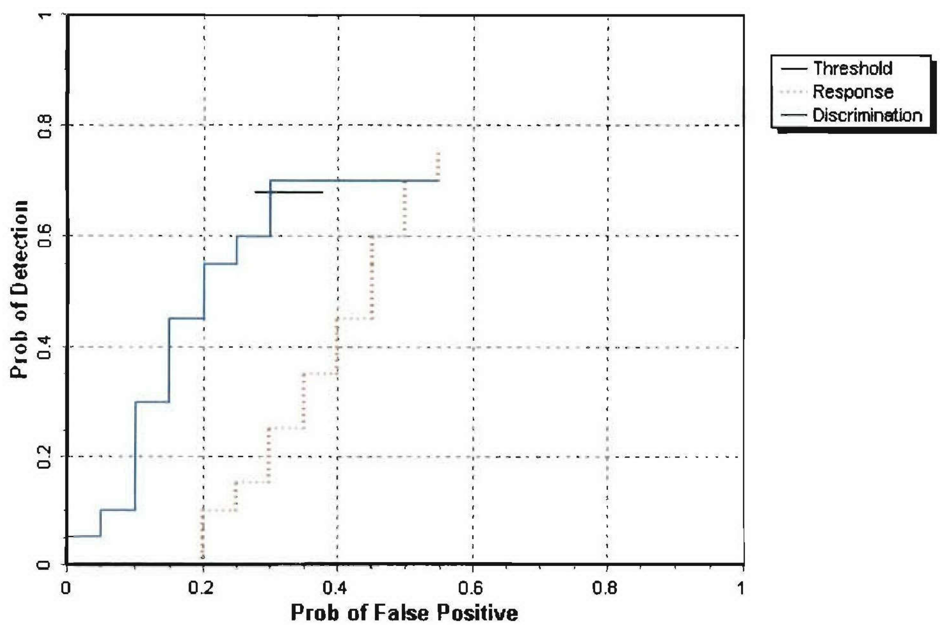

Figure 4. MTADS MAG/towed open field probability of detection for response and discrimination stages versus their respective probability of false positive for all ordnance larger than 20 mm.

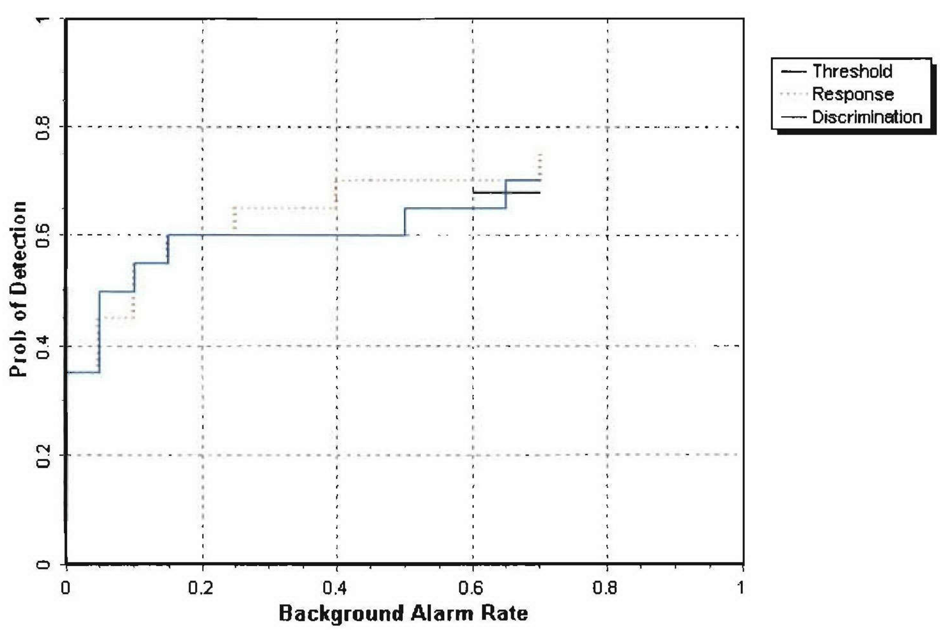

Figure 5. MTADS MAG/towed open field probability of detection for response and discrimination stages versus their respective background alarm rate for all ordnance larger than 20 mm.

4.3 PERFORMANCE SUMMARIES

Results for the Open Field test, broken out by size, depth and nonstandard ordnance, are presented in Tables 5a and 5b (for cost results, see section 5). Results by size and depth include both standard and nonstandard ordnance. The results by size show how well the demonstrator did at detecting/discriminating ordnance of a certain caliber range (see app A for size definitions). The results are relative to the number of ordnances emplaced. Depth is measured from the geometric center of anomalies.

The RESPONSE STAGE results are derived from the list of anomalies above the demonstrator-provided noise level. The results for the DISCRIMINATION STAGE are derived from the demonstrator's recommended threshold for optimizing UXO field cleanup by minimizing false digs and maximizing ordnance recovery. The lower 90-percent confidence limit on probability of detection and probability of false positive was calculated assuming that the number of detections and false positives are binomially distributed random variables. All results in Tables 5a and 5b have been rounded to protect the ground truth. However, lower confidence limits were calculated using actual results.

The overall ground truth is composed of ferrous and non-ferrous anomalies. Due to limitations of the magnetometer, the non-ferrous items cannot be detected. Therefore, the summary presented in Table 5a exhibits results based on the subset of the ground truth that is solely the ferrous anomalies. Table 5b exhibits results based on the full ground truth. All other tables presented in this section are based on scoring against the ferrous only ground truth. The response stage noise level and recommended discrimination stage threshold values are provided by the demonstrator.

TABLE 5a. SUMMARY OF OPEN FIELD RESULTS (FERROUS ONLY)

Metric	Overall	Standard	Nonstandard	By Size			By Depth, m		
				Small	Medium	Large	< 0.3	0.3 to <1	>= 1
RESPONSE STAGE									
P_d	0.70	0.70	0.60	0.60	0.65	0.85	0.75	0.65	0.55
P_d Low 90% Conf	0.64	0.67	0.55	0.52	0.60	0.77	0.68	0.60	0.48
P_d Upper 90% Conf	0.71	0.77	0.67	0.65	0.72	0.89	0.79	0.72	0.66
P_{fp}	0.55	-	-	-	-	-	0.50	0.60	0.75
P_{fp} Low 90% Conf	0.52	-	-	-	-	-	0.45	0.55	0.56
P_{fp} Upper 90% Conf	0.56	-	-	-	-	-	0.51	0.61	0.89
BAR	0.70	-	-	-	-	-	-	-	-
DISCRIMINATION STAGE									
P_d	0.65	0.70	0.55	0.50	0.65	0.80	0.70	0.65	0.55
P_d Low 90% Conf	0.59	0.63	0.49	0.46	0.56	0.72	0.62	0.56	0.45
P_d Upper 90% Conf	0.67	0.73	0.61	0.59	0.68	0.86	0.73	0.69	0.63
P_{fp}	0.35	-	-	-	-	-	0.35	0.30	0.40
P_{fp} Low 90% Conf	0.31	-	-	-	-	-	0.32	0.28	0.21
P_{fp} Upper 90% Conf	0.35	-	-	-	-	-	0.38	0.34	0.57
BAR	0.65	-	-	-	-	-	-	-	-

Response Stage Noise Level: 19.00
Recommended Discrimination Stage Threshold: 490.50

TABLE 5b. SUMMARY OF OPEN FIELD RESULTS (FULL GROUND TRUTH)

Metric	Overall	Standard	Nonstandard	By Size			By Depth, m		
				Small	Medium	Large	< 0.3	0.3 to <1	>= 1
RESPONSE STAGE									
P_d	0.60	0.65	0.55	0.45	0.65	0.85	0.60	0.60	0.55
P_d Low 90% Conf	0.56	0.58	0.49	0.38	0.60	0.77	0.55	0.55	0.47
P_d Upper 90% Conf	0.63	0.67	0.61	0.49	0.72	0.89	0.65	0.67	0.64
P_{fp}	0.50	-	-	-	-	-	0.45	0.55	0.75
P_{fp} Low 90% Conf	0.50	-	-	-	-	-	0.44	0.54	0.56
P_{fp} Upper 90% Conf	0.54	-	-	-	-	-	0.50	0.60	0.89
BAR	0.70	-	-	-	-	-	-	-	-
DISCRIMINATION STAGE									
P_d	0.55	0.60	0.50	0.40	0.65	0.80	0.55	0.60	0.50
P_d Low 90% Conf	0.52	0.55	0.44	0.33	0.56	0.72	0.50	0.52	0.44
P_d Upper 90% Conf	0.59	0.64	0.56	0.44	0.68	0.86	0.61	0.64	0.61
P_{fp}	0.30	-	-	-	-	-	0.35	0.30	0.40
P_{fp} Low 90% Conf	0.30	-	-	-	-	-	0.31	0.27	0.21
P_{fp} Upper 90% Conf	0.34	-	-	-	-	-	0.37	0.33	0.57
BAR	0.65	-	-	-	-	-	-	-	-

Response Stage Noise Level: 19.00
Recommended Discrimination Stage Threshold 490.50

Note: The recommended discrimination stage threshold values are provided by the demonstrator.

4.4 EFFICIENCY, REJECTION RATES, AND TYPE CLASSIFICATION

Efficiency and rejection rates are calculated to quantify the discrimination ability at specific points of interest on the ROC curve: (1) at the point where no decrease in P_d is suffered (i.e., the efficiency is by definition equal to one) and (2) at the operator selected threshold. These values are reported in Table 6.

TABLE 6. EFFICIENCY AND REJECTION RATES

	Efficiency (E)	False Positive Rejection Rate	Background Alarm Rejection Rate
At Operating Point	0.93	0.39	0.07
With No Loss of P_d	1.00	0.18	0.00

At the demonstrator's recommended setting, the ordnance items that were detected and correctly discriminated were further scored on whether their correct type could be identified (table 7). Correct type examples include "20-mm projectile, 105-mm HEAT Projectile, and 2.75-inch Rocket". A list of the standard type declaration required for each ordnance item was provided to demonstrators prior to testing. For example, the standard type for the three example items are 20mmP, 105H, and 2.75in, respectively.

TABLE 7. CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS UXO

Size	Percentage Correct
Small	27.3
Medium	22.1
Large	34.5
Overall	27.4

4.5 LOCATION ACCURACY

The mean location error and standard deviations appear in Table 8. These calculations are based on average missed depth for ordnance correctly identified in the discrimination stage. Depths are measured from the closest point of the ordnance to the surface. For the Blind Grid, only depth errors are calculated, since (X, Y) positions are known to be the centers of each grid square.

TABLE 8. MEAN LOCATION ERROR AND STANDARD DEVIATION (M)

	Mean	Standard Deviation
Northing	-0.01	0.14
Easting	-0.03	0.14
Depth	-0.08	0.26

SECTION 5. ON-SITE LABOR COSTS

A standardized estimate for labor costs associated with this effort was calculated as follows: the first person at the test site was designated "supervisor", the second person was designated "data analyst", and the third and following personnel were considered "field support". Standardized hourly labor rates were charged by title: supervisor at $95.00/hour, data analyst at $57.00/hour, and field support at $28.50/hour.

Government representatives monitored on-site activity. All on-site activities were grouped into one of ten categories: initial setup/mobilization, daily setup/stop, calibration, collecting data, downtime due to break/lunch, downtime due to equipment failure, downtime due to equipment/data checks or maintenance, downtime due to weather, downtime due to demonstration site issue, or demobilization. See Appendix D for the daily activity log. See section 3.4 for a summary of field activities.

The standardized cost estimate associated with the labor needed to perform the field activities is presented in Table 9. Note that calibration time includes time spent in the Calibration Lanes as well as field calibrations. "Site survey time" includes daily setup/stop time, collecting data, breaks/lunch, downtime due to equipment/data checks or maintenance, downtime due to failure, and downtime due to weather.

TABLE 9. ON-SITE LABOR COSTS

	No. People	Hourly Wage	Hours	Cost
Initial Setup				
Supervisor	1	$95.00	1.50	$142.50
Data Analyst	1	57.00	1.50	85.50
Field Support	1	28.50	1.50	42.75
SubTotal				**$270.75**
Calibration				
Supervisor	1	$95.00	0.75	$71.25
Data Analyst	1	57.00	0.75	42.75
Field Support	1	28.50	0.75	21.38
SubTotal				**$135.38**
Site Survey				
Supervisor	1	$95.00	7.23	$686.85
Data Analyst	1	57.00	7.23	412.11
Field Support	1	28.50	7.23	206.06
SubTotal				**$1,305.02**

See notes at end of table.

TABLE 9 (CONT'D)

	No. People	Hourly Wage	Hours	Cost
Demobilization				
Supervisor	1	$95.00	1.0	$95.00
Data Analyst	1	57.00	1.0	57.00
Field Support	1	28.50	1.0	28.50
Subtotal				**$180.50**
Total				**$1,891.65**

Notes: Calibration time includes time spent in the Calibration Lanes as well as calibration before each data run.

Site Survey time includes daily setup/stop time, collecting data, breaks/lunch, downtime due to system maintenance, failure, and weather.

SECTION 6. COMPARISON OF RESULTS TO BLIND GRID DEMONSTRATION (BASED ON FERROUS ONLY GROUND TRUTH)

6.1 SUMMARY OF RESULTS FROM BLIND GRID DEMONSTRATION

Table 10 shows the results from the Blind Grid survey conducted prior to surveying the Open Field during the same site visit in June of 2004. Due to the system utilizing magnetometer type sensors, all results presented in the following section have been based on performance scoring against the ferrous only ground truth anomalies. For more details on the Blind Grid survey results reference section 2.1.6.

TABLE 10. SUMMARY OF BLIND GRID RESULTS FOR THE MTADS MAG/TOWED (FERROUS ONLY)

Metric	Overall	Standard	Nonstandard	By Size			By Depth, m		
				Small	Medium	Large	< 0.3	0.3 to <1	>= 1
RESPONSE STAGE									
P_d	0.70	0.80	0.50	0.65	0.70	0.90	0.80	0.80	0.30
P_d Low 90% Conf	0.62	0.73	0.34	0.52	0.55	0.66	0.67	0.65	0.14
P_d Upper 90% Conf	0.77	0.89	0.62	0.77	0.79	0.99	0.89	0.88	0.52
P_{fp}	0.85	-	-	-	-	-	0.85	0.80	1.00
P_{fp} Low 90% Conf	0.78	-	-	-	-	-	0.76	0.71	0.63
P_{fp} Upper 90% Conf	0.89	-	-	-	-	-	0.92	0.89	1.00
P_{ba}	0.10	-	-	-	-	-	-	-	-
DISCRIMINATION STAGE									
P_d	0.45	0.60	0.30	0.45	0.50	0.40	0.55	0.50	0.25
P_d Low 90% Conf	0.39	0.47	0.16	0.32	0.39	0.19	0.42	0.36	0.09
P_d Upper 90% Conf	0.55	0.68	0.43	0.58	0.64	0.65	0.68	0.64	0.44
P_{fp}	0.50	-	-	-	-	-	0.55	0.50	0.40
P_{fp} Low 90% Conf	0.45	-	-	-	-	-	0.45	0.39	0.11
P_{fp} Upper 90% Conf	0.59	-	-	-	-	-	0.65	0.61	0.75
P_{ba}	0.05	-	-	-	-	-	-	-	-

6.2 COMPARISON OF ROC CURVES USING ALL ORDNANCE CATEGORIES

Figure 6 shows P_d^{res} versus the respective P_{fp} over all ordnance categories. Figure 7 shows P_d^{disc} versus their respective P_{fp} over all ordnance categories. Figure 7 uses horizontal lines to illustrate the performance of the demonstrator at the recommended discrimination threshold levels, defining the subset of targets the demonstrator would recommend digging based on discrimination. The ROC curves in this section are a sole reflection of the ferrous only survey.

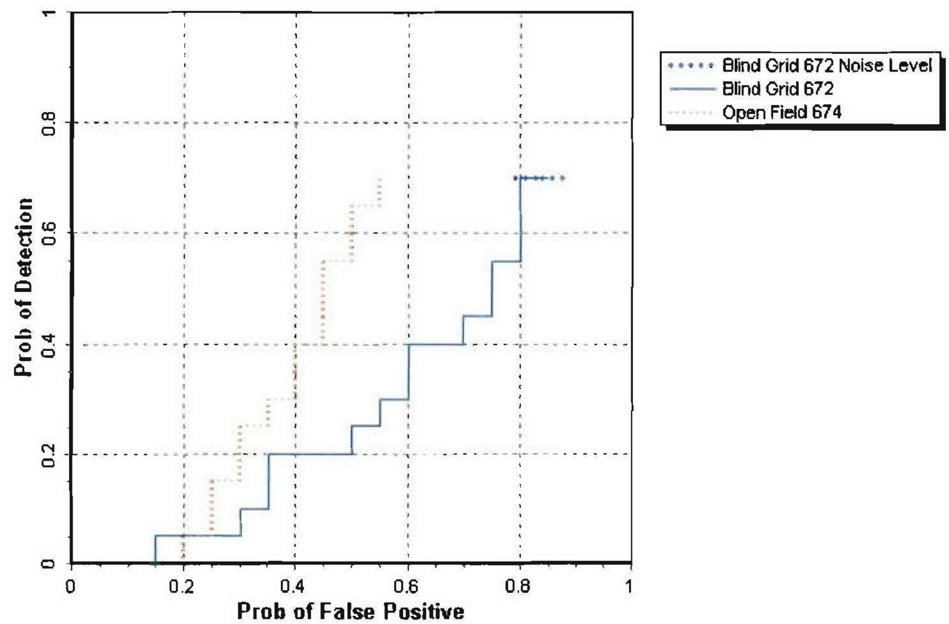

Figure 6. MTADS MAG/towed P_d^{res} stages versus the respective P_{fp} over all ordnance categories combined.

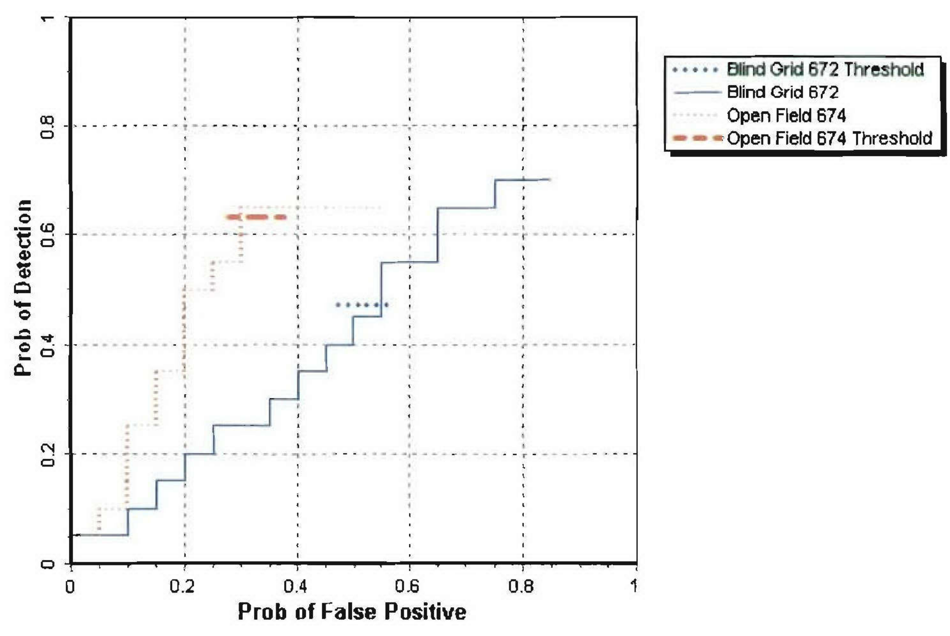

Figure 7. MTADS MAG/towed P_d^{disc} versus the respective P_{fp} over all ordnance categories combined.

6.3 COMPARISON OF ROC CURVES USING ORDNANCE LARGER THAN 20 MM

Figure 8 shows the P_d^{res} versus the respective probability of P_{fp} over ordnance larger than 20 mm. Figure 9 shows P_d^{disc} versus the respective P_{fp} over ordnance larger than 20 mm. Figure 9 uses horizontal lines to illustrate the performance of the demonstrator at the recommended discrimination threshold levels, defining the subset of targets the demonstrator would recommend digging based on discrimination.

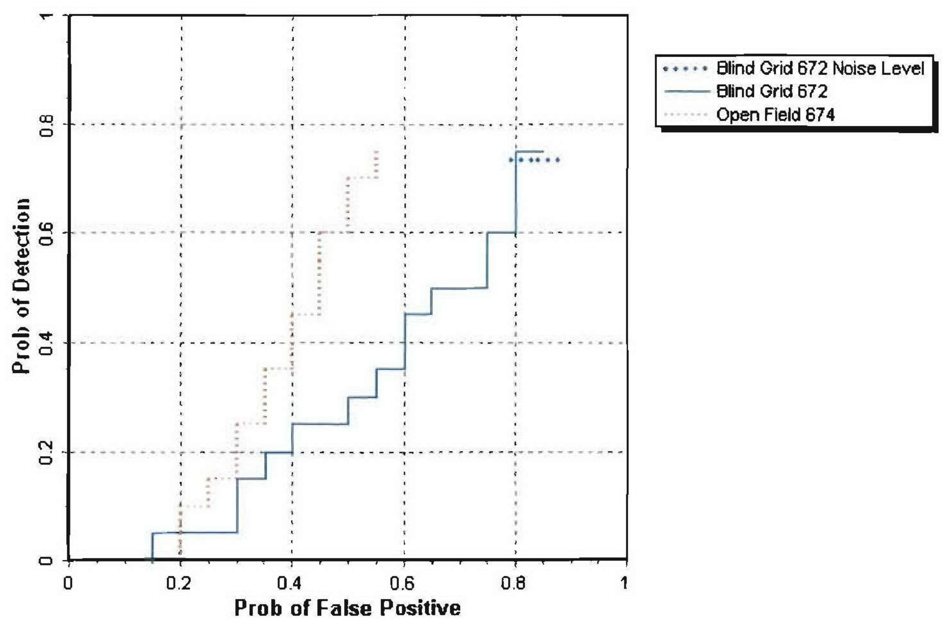

Figure 8. MTADS MAG/towed P_d^{res} versus the respective P_{fp} for ordnance larger than 20 mm.

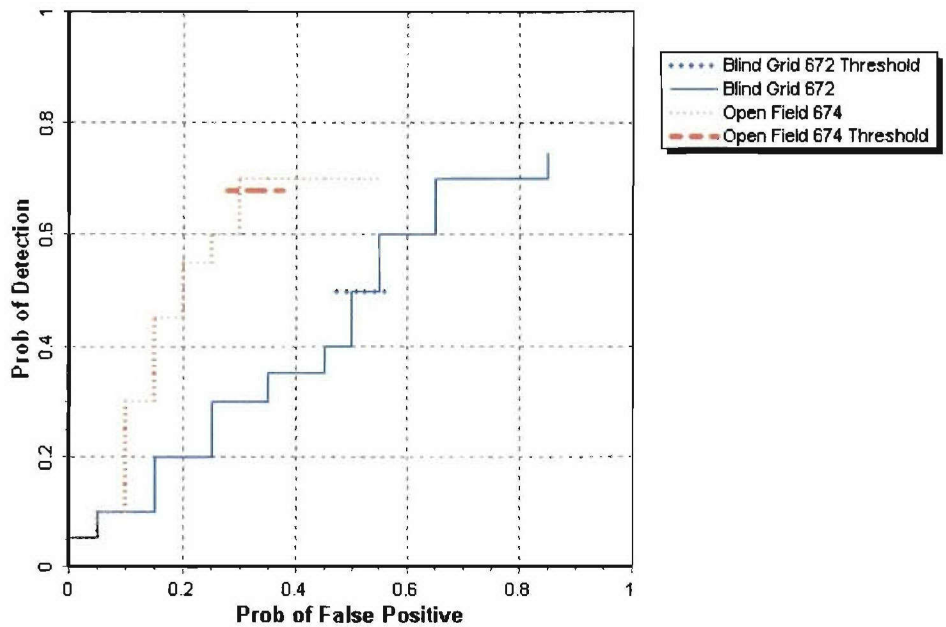

Figure 9. MTADS MAG/towed P_d^{disc} versus the respective P_{fp} for ordnance larger than 20 mm.

6.4 STATISTICAL COMPARISONS

Statistical Chi-square significance tests were used to compare results between the Blind Grid and Open Field scenarios. The intent of the comparison is to determine if the feature introduced in each scenario has a degrading effect on the performance of the sensor system. However, any modifications in the UXO sensor system during the test, like changes in the processing or changes in the selection of the operating threshold, will also contribute to performance differences.

The Chi-square test for comparison between ratios was used at a significance level of 0.05 to compare Blind Grid to Open Field with regard to P_d^{res}, P_d^{disc}, P_{fp}^{res} and P_{fp}^{disc}, Efficiency and Rejection Rate. These results are presented in Table 11. A detailed explanation and example of the Chi-square application is located in Appendix A.

TABLE 11. CHI-SQUARE RESULTS - BLIND GRID VERSUS OPEN FIELD

Metric	Small	Medium	Large	Overall
P_d^{res}	Not Significant	Not Significant	Not Significant	Not Significant
P_d^{disc}	Not Significant	Not Significant	Significant	Significant
P_{fp}^{res}	Not Significant	Not Significant	Not Significant	Significant
P_{fp}^{disc}	-	-	-	Significant
Efficiency	-			Significant
Rejection rate	-	-	-	Not Significant

SECTION 7. APPENDIXES

APPENDIX A. TERMS AND DEFINITIONS

GENERAL DEFINITIONS

Anomaly: Location of a system response deemed to warrant further investigation by the demonstrator for consideration as an emplaced ordnance item.

Detection: An anomaly location that is within R_{halo} of an emplaced ordnance item.

Emplaced Ordnance: An ordnance item buried by the government at a specified location in the test site.

Emplaced Clutter: A clutter item (i.e., non-ordnance item) buried by the government at a specified location in the test site.

R_{halo}: A pre-determined radius about the periphery of an emplaced item (clutter or ordnance) within which a location identified by the demonstrator as being of interest is considered to be a response from that item. If multiple declarations lie within R_{halo} of any item (clutter or ordnance), the declaration with the highest signal output within the R_{halo} will be utilized. For the purpose of this program, a circular halo 0.5 meters in radius will be placed around the center of the object for all clutter and ordnance items less than 0.6 meters in length. When ordnance items are longer than 0.6 meters, the halo becomes an ellipse where the minor axis remains 1 meter and the major axis is equal to the length of the ordnance plus 1 meter.

Small Ordnance: Caliber of ordnance less than or equal to 40 mm (includes 20-mm projectile, 40-mm projectile, submunitions BLU-26, BLU-63, and M42).

Medium Ordnance: Caliber of ordnance greater than 40 mm and less than or equal to 81 mm (includes 57-mm projectile, 60-mm mortar, 2.75 in. Rocket, MK118 Rockeye, 81-mm mortar).

Large Ordnance: Caliber of ordnance greater than 81 mm (includes 105-mm HEAT, 105-mm projectile, 155-mm projectile, 500-pound bomb).

Shallow: Items buried less than 0.3 meter below ground surface.

Medium: Items buried greater than or equal to 0.3 meter and less than 1 meter below ground surface.

Deep: Items buried greater than or equal to 1 meter below ground surface.

Response Stage Noise Level: The level that represents the point below which anomalies are not considered detectable. Demonstrators are required to provide the recommended noise level for the Blind Grid test area.

Discrimination Stage Threshold: The demonstrator selected threshold level that they believe provides optimum performance of the system by retaining all detectable ordnance and rejecting the maximum amount of clutter. This level defines the subset of anomalies the demonstrator would recommend digging based on discrimination.

Binomially Distributed Random Variable: A random variable of the type which has only two possible outcomes, say success and failure, is repeated for n independent trials with the probability p of success and the probability 1-p of failure being the same for each trial. The number of successes x observed in the n trials is an estimate of p and is considered to be a binomially distributed random variable.

RESPONSE AND DISCRIMINATION STAGE DATA

The scoring of the demonstrator's performance is conducted in two stages. These two stages are termed the RESPONSE STAGE and DISCRIMINATION STAGE. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of false positive (P_{fp}) and those that do not correspond to any known item, termed background alarms.

The RESPONSE STAGE scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate ordnance from other anomalies. For the RESPONSE STAGE, the demonstrator provides the scoring committee with the location and signal strength of all anomalies that the demonstrator has deemed sufficient to warrant further investigation and/or processing as potential emplaced ordnance items. This list is generated with minimal processing (e.g., this list will include all signals above the system noise threshold). As such, it represents the most inclusive list of anomalies.

The DISCRIMINATION STAGE evaluates the demonstrator's ability to correctly identify ordnance as such, and to reject clutter. For the same locations as in the RESPONSE STAGE anomaly list, the DISCRIMINATION STAGE list contains the output of the algorithms applied in the discrimination-stage processing. This list is prioritized based on the demonstrator's determination that an anomaly location is likely to contain ordnance. Thus, higher output values are indicative of higher confidence that an ordnance item is present at the specified location. For electronic signal processing, priority ranking is based on algorithm output. For other systems, priority ranking is based on human judgment. The demonstrator also selects the threshold that the demonstrator believes will provide "optimum" system performance, (i.e., that retains all the detected ordnance and rejects the maximum amount of clutter).

Note: The two lists provided by the demonstrator contain identical numbers of potential target locations. They differ only in the priority ranking of the declarations.

RESPONSE STAGE DEFINITIONS

Response Stage Probability of Detection (P_d^{res}): P_d^{res} = (No. of response-stage detections)/(No. of emplaced ordnance in the test site).

Response Stage False Positive (fp^{res}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Response Stage Probability of False Positive (P_{fp}^{res}): P_{fp}^{res} = (No. of response-stage false positives)/(No. of emplaced clutter items).

Response Stage Background Alarm (ba^{res}): An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced ordnance or emplaced clutter item.

Response Stage Probability of Background Alarm (P_{ba}^{res}): Blind Grid only: P_{ba}^{res} = (No. of response-stage background alarms)/(No. of empty grid locations).

Response Stage Background Alarm Rate (BAR^{res}): Open Field only: BAR^{res} = (No. of response-stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{res}, P_{fp}^{res}, P_{ba}^{res}, and BAR^{res} are functions of t^{res}, the threshold applied to the response-stage signal strength. These quantities can therefore be written as $P_d^{res}(t^{res})$, $P_{fp}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to response-stage data that discriminates ordnance from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to ordnance, as well as those that the demonstrator has high confidence correspond to nonordnance or background returns. The former should be ranked with highest priority and the latter with lowest.

Discrimination Stage Probability of Detection (P_d^{disc}): P_d^{disc} = (No. of discrimination-stage detections)/(No. of emplaced ordnance in the test site).

Discrimination Stage False Positive (fp^{disc}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Discrimination Stage Probability of False Positive (P_{fp}^{disc}): P_{fp}^{disc} = (No. of discrimination stage false positives)/(No. of emplaced clutter items).

Discrimination Stage Background Alarm (ba^{disc}): An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced ordnance or emplaced clutter item.

Discrimination Stage Probability of Background Alarm (P_{ba}^{disc}): P_{ba}^{disc} = (No. of discrimination-stage background alarms)/(No. of empty grid locations).

Discrimination Stage Background Alarm Rate (BAR^{disc}): BAR^{disc} = (No. of discrimination-stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{disc}, P_{fp}^{disc}, P_{ba}^{disc}, and BAR^{disc} are functions of t^{disc}, the threshold applied to the discrimination-stage signal strength. These quantities can therefore be written as $P_d^{disc}(t^{disc})$, $P_{fp}^{disc}(t^{disc})$, $P_{ba}^{disc}(t^{disc})$, and $BAR^{disc}(t^{disc})$.

RECEIVER-OPERATING CHARACERISTIC (ROC) CURVES

ROC curves at both the response and discrimination stages can be constructed based on the above definitions. The ROC curves plot the relationship between P_d versus P_{fp} and P_d versus BAR or P_{ba} as the threshold applied to the signal strength is varied from its minimum (t_{min}) to its maximum (t_{max}) value.[1] Figure A-1 shows how P_d versus P_{fp} and P_d versus BAR are combined into ROC curves. Note that the "res" and "disc" superscripts have been suppressed from all the variables for clarity.

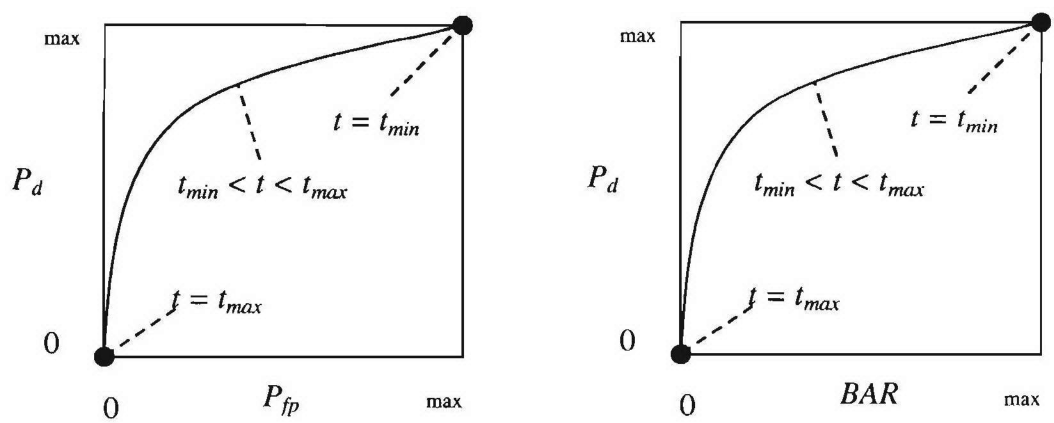

Figure A-1. ROC curves for open field testing. Each curve applies to both the response and discrimination stages.

---

[1]Strictly speaking, ROC curves plot the P_d versus P_{ba} over a pre-determined and fixed number of detection opportunities (some of the opportunities are located over ordnance and others are located over clutter or blank spots). In an open field scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the open field ROC curves do not have information from low signal-output locations, and, furthermore, different contractors report their signals over a different set of locations on the ground. These ROC curves are thus not true to the strict definition of ROC curves as defined in textbooks on detection theory. Note, however, that the ROC curves obtained in the Blind Grid test sites are true ROC curves.

METRICS TO CHARACTERIZE THE DISCRIMINATION STAGE

The demonstrator is also scored on efficiency and rejection ratio, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of ordnance detections from the anomaly list, while rejecting the maximum number of anomalies arising from nonordnance items. The efficiency measures the amount of detected ordnance retained by the discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the entire response list, i.e., the maximum ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate.

Efficiency (E): $E = P_d^{disc}(t^{disc})/P_d^{res}(t_{min}^{res})$; Measures (at a threshold of interest), the degree to which the maximum theoretical detection performance of the sensor system (as determined by the response stage tmin) is preserved after application of discrimination techniques. Efficiency is a number between 0 and 1. An efficiency of 1 implies that all of the ordnance initially detected in the response stage was retained at the specified threshold in the discrimination stage, t^{disc}.

False Positive Rejection Rate (R_{fp}): $R_{fp} = 1 - [P_{fp}^{disc}(t^{disc})/P_{fp}^{res}(t_{min}^{res})]$; Measures (at a threshold of interest), the degree to which the sensor system's false positive performance is improved over the maximum false positive performance (as determined by the response stage tmin). The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all emplaced clutter initially detected in the response stage were correctly rejected at the specified threshold in the discrimination stage.

Background Alarm Rejection Rate (R_{ba}):

Blind Grid: $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$.
Open Field: $R_{ba} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})])$.

Measures the degree to which the discrimination stage correctly rejects background alarms initially detected in the response stage. The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all background alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

CHI-SQUARE COMPARISON EXPLANATION:

The Chi-square test for differences in probabilities (or 2 x 2 contingency table) is used to analyze two samples drawn from two different populations to see if both populations have the same or different proportions of elements in a certain category. More specifically, two random samples are drawn, one from each population, to test the null hypothesis that the probability of event A (some specified event) is the same for both populations (ref 3).

A 2 x 2 contingency table is used in the Standardized UXO Technology Demonstration Site Program to determine if there is reason to believe that the proportion of ordnance correctly detected/discriminated by demonstrator X's system is significantly degraded by the more challenging terrain feature introduced. The test statistic of the 2 x 2 contingency table is the

Chi-square distribution with one degree of freedom. Since an association between the more challenging terrain feature and relatively degraded performance is sought, a one-sided test is performed. A significance level of 0.05 is chosen which sets a critical decision limit of 2.71 from the Chi-square distribution with one degree of freedom. It is a critical decision limit because if the test statistic calculated from the data exceeds this value, the two proportions tested will be considered significantly different. If the test statistic calculated from the data is less than this value, the two proportions tested will be considered not significantly different.

An exception must be applied when either a 0 or 100 percent success rate occurs in the sample data. The Chi-square test cannot be used in these instances. Instead, Fischer's test is used and the critical decision limit for one-sided tests is the chosen significance level, which in this case is 0.05. With Fischer's test, if the test statistic is less than the critical value, the proportions are considered to be significantly different.

Standardized UXO Technology Demonstration Site examples, where blind grid results are compared to those from the open field and open field results are compared to those from one of the scenarios, follow. It should be noted that a significant result does not prove a cause and effect relationship exists between the two populations of interest; however, it does serve as a tool to indicate that one data set has experienced a degradation in system performance at a large enough level than can be accounted for merely by chance or random variation. Note also that a result that is not significant indicates that there is not enough evidence to declare that anything more than chance or random variation within the same population is at work between the two data sets being compared.

Demonstrator X achieves the following overall results after surveying each of the three progressively more difficult areas using the same system (results indicate the number of ordnance detected divided by the number of ordnance emplaced):

	Blind Grid	Open Field	Moguls
P_d^{res}	100/100 = 1.0	8/10 = .80	20/33 = .61
P_d^{disc}	80/100 = 0.80	6/10 = .60	8/33 = .24

P_d^{res}: BLIND GRID versus OPEN FIELD. Using the example data above to compare probabilities of detection in the response stage, all 100 ordnance out of 100 emplaced ordnance items were detected in the blind grid while 8 ordnance out of 10 emplaced were detected in the open field. Fischer's test must be used since a 100 percent success rate occurs in the data. Fischer's test uses the four input values to calculate a test statistic of 0.0075 that is compared against the critical value of 0.05. Since the test statistic is less than the critical value, the smaller response stage detection rate (0.80) is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the detection ability of demonstrator X's system seems to have been degraded in the open field relative to results from the blind grid using the same system.

P_d^{disc}: BLIND GRID versus OPEN FIELD. Using the example data above to compare probabilities of detection in the discrimination stage, 80 out of 100 emplaced ordnance items were correctly discriminated as ordnance in blind grid testing while 6 ordnance out of 10 emplaced were correctly discriminated as such in open field-testing. Those four values are used to calculate a test statistic of 1.12. Since the test statistic is less than the critical value of 2.71, the two discrimination stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{res}: OPEN FIELD versus MOGULS. Using the example data above to compare probabilities of detection in the response stage, 8 out of 10 and 20 out of 33 are used to calculate a test statistic of 0.56. Since the test statistic is less than the critical value of 2.71, the two response stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{disc}: OPEN FIELD versus MOGULS. Using the example data above to compare probabilities of detection in the discrimination stage, 6 out of 10 and 8 out of 33 are used to calculate a test statistic of 2.98. Since the test statistic is greater than the critical value of 2.71, the smaller discrimination stage detection rate is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the ability of demonstrator X to correctly discriminate seems to have been degraded by the mogul terrain relative to results from the flat open field using the same system.

APPENDIX B. DAILY WEATHER LOGS

TABLE B-1. WEATHER LOG

Date & Time	Average Temp (°F)	Maximum Temp (°F)	Minimum Temp (°F)	Total Precip (in.)
06/21/2004 00:00:00	57	57.9	56.2	0
06/21/2004 01:00:00	56	56.8	54.6	0
06/21/2004 02:00:00	55	56	53.9	0
06/21/2004 03:00:00	54	54.6	53.4	0
06/21/2004 04:00:00	54.1	54.7	53.5	0
06/21/2004 05:00:00	54.1	54.8	53.3	0
06/21/2004 06:00:00	56.2	59	53.5	0
06/21/2004 07:00:00	62.8	65.7	58.6	0
06/21/2004 08:00:00	68.7	70.8	65.2	0
06/21/2004 09:00:00	71.5	72.9	70	0
06/21/2004 10:00:00	73.2	74.9	71.2	0
06/21/2004 11:00:00	74.6	76.3	73.8	0
06/21/2004 12:00:00	75.5	76.7	74.2	0
06/21/2004 13:00:00	77.1	78.1	76.2	0
06/21/2004 14:00:00	77.9	79.1	76.9	0
06/21/2004 15:00:00	78	78.9	77.2	0
06/21/2004 16:00:00	78.3	78.9	77.5	0
06/21/2004 17:00:00	77.9	78.8	77.3	0
06/21/2004 18:00:00	77.2	77.9	76.3	0
06/21/2004 19:00:00	75.7	76.5	74.6	0
06/21/2004 20:00:00	73.6	75	72.7	0
06/21/2004 21:00:00	73.5	73.9	73	0
06/21/2004 22:00:00	73.4	74.5	72.6	0

TABLE B-1. (CONT'D)

Date & Time	Average Temp (°F)	Maximum Temp (°F)	Minimum Temp (°F)	Total Precip (in.)
06/21/2004 23:00:00	73.2	74.3	71.5	0
06/22/2004 00:00:00	70.7	71.9	69.6	0
06/22/2004 01:00:00	68.9	70	68.2	0
06/22/2004 02:00:00	68.9	69.4	68.2	0
06/22/2004 03:00:00	69	73.1	67.6	0
06/22/2004 04:00:00	73.7	74.2	73	0
06/22/2004 05:00:00	73.6	74	73	0
06/22/2004 06:00:00	73.3	74	72.9	0
06/22/2004 07:00:00	74.7	75.6	73.7	0
06/22/2004 08:00:00	76	77.3	75	0
06/22/2004 09:00:00	76.4	77.3	75.6	0
06/22/2004 10:00:00	77.6	79.2	76.1	0
06/22/2004 11:00:00	78.9	80.3	77.5	0
06/22/2004 12:00:00	80.2	81.8	79.4	0
06/22/2004 13:00:00	81.1	82.7	80	0
06/22/2004 14:00:00	83	83.8	82.1	0
06/22/2004 15:00:00	84.1	86.1	82.7	0
06/22/2004 16:00:00	83.4	84.9	82.5	0
06/22/2004 17:00:00	82.2	82.9	80.5	0
06/22/2004 18:00:00	78.5	81.1	72.4	0.1
06/22/2004 19:00:00	71.8	72.6	70.6	0.14
06/22/2004 20:00:00	70.8	71.9	69.9	0
06/22/2004 21:00:00	69.8	70.4	69.3	0
06/22/2004 22:00:00	69.8	70.5	69	0
06/22/2004 23:00:00	69.9	71	68.8	0

APPENDIX C. SOIL MOISTURE

Date 6/21/04
Times: 0800 hours, 1600 hours

Probe Location:	Layer, in.	AM Reading, %	PM Reading, %
Wet Area	0 to 6	60.3	60.2
	6 to 12	75.3	75.7
	12 to 24	77.8	78.3
	24 to 36	55.3	55.2
	36 to 48	48.3	48.7
Wooded Area	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		
Open Area	0 to 6	21.3	21.2
	6 to 12	6.3	6.5
	12 to 24	18.2	18.1
	24 to 36	26.3	26.5
	36 to 48	54.1	54.3
Calibration Lanes	0 to 6	40.3	40.2
	6 to 12	38.4	38.5
	12 to 24	1.5	1.8
	24 to 36	4.2	4.0
	36 to 48	4.3	4.5
Blind Grid/Moguls	0 to 6	3.9	4.0
	6 to 12	25.7	26.1
	12 to 24	35.9	36.3
	24 to 36	37.8	37.5
	36 to 48	38.5	38.4

Date 6/22/04
Times: 0800 hours

Probe Location:	Layer, in.	AM Reading, %	PM Reading, %
Wet Area	0 to 6	60.1	
	6 to 12	75.9	
	12 to 24	78.2	
	24 to 36	55.7	
	36 to 48	48.3	
Wooded Area	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		
Open Area	0 to 6	21.1	
	6 to 12	6.8	
	12 to 24	18.5	
	24 to 36	26.1	
	36 to 48	53.9	
Calibration Lanes	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		
Blind Grid/Moguls	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		

APPENDIX D. DAILY ACTIVITY LOGS

Date	No. of People	Area Tested	Status Start Time	Status Stop Time	Duration, min	Operational Status	OP Stat Code	Operational Status - Comments	Track Method	Track Method=Other Explain	Pattern	Field Conditions	
6/21/2004	3	OPEN FIELD	745	915	90	INITIAL MOBILIZATION	1	INITIAL MOBILIZATION	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	915	1015	60	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1015	1020	5	DOWNTIME MAINTENANCE CHECK	7	DOWNLOAD DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	1/2 OPEN FIELD, 1/4 BG, 1/4 CAL LANE	1020	1105	45	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	1/2 OPEN FIELD, 1/4 BG, 1/4 CAL LANE	1105	1110	5	DOWNTIME MAINTENANCE CHECK	7	DOWNLOAD DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	1/2 OPEN FIELD, ¼ BG, 1/4 CAL LANE	1110	1215	65	LUNCH/BREAK	5	LUNCH/BREAK	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	1/2 OPEN FIELD, ¼ BG, 1/4 CAL LANE	1215	1320	65	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1320	1325	5	DOWNTIME MAINTENANCE CHECK	7	DOWNLOAD DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1325	1430	65	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1430	1435	5	DOWNTIME MAINTENANCE CHECK	7	DOWNLOAD DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1435	1500	25	LUNCH/BREAK	5	LUNCH/BREAK	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1500	1600	60	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY

Note: Activities pertinent to this specific demonstration are indicated in highlighted text.

Date	No. of People	Area Tested	Status Start Time	Status Stop Time	Duration, min	Operational Status	OP Stat Code	Operational Status - Comments	Track Method	Track Method=Other Explain	Pattern	Field Conditions	
6/21/2004	3	OPEN FIELD	1600	1605	5	DOWNTIME MAINTENANCE CHECK	7	DOWNLOAD DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1605	1630	25	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1630	1645	15	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/21/2004	3	OPEN FIELD	1645	1710	25	DAILY START STOP	3	END OF DAILY OPERATIONS, BREAKDOWN	GPS	NA	LINEAR	SUNNY	MUDDY
6/22/2004	3	OPEN FIELD	740	800	20	DAILY START STOP	3	SET UP BEGIN OPERATIONS	GPS	NA	LINEAR	SUNNY	MUDDY
6/22/2004	3	OPEN FIELD	800	830	30	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/22/2004	3	OPEN FIELD	830	835	5	DOWNTIME MAINTENANCE CHECK	7	DOWNLOAD DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/22/2004	3	OPEN FIELD	835	930	55	COLLECT DATA	4	COLLECT DATA	GPS	NA	LINEAR	SUNNY	MUDDY
6/22/2004	3	OPEN FIELD	930	1030	60	DEMOBILIZATION	10	DEMOBILIZATION	GPS	NA	LINEAR	SUNNY	MUDDY

Note: Activities pertinent to this specific demonstration are indicated in highlighted text.

APPENDIX E. REFERENCES

1. Standardized UXO Technology Demonstration Site Handbook, DTC Project No. 8-CO-160-000-473, Report No. ATC-8349, March 2002.

2. Aberdeen Proving Ground Soil Survey Report, October 1998.

3. Data Summary, UXO Standardized Test Site: APG Soils Description, May 2002.

4. Yuma Proving Ground Soil Survey Report, May 2003.

APPENDIX F. ABBREVIATIONS

AEC = U.S. Army Environmental Center
APG = Aberdeen Proving Ground
ASCII = American Standard Code for Information Interchange.
ATC = U.S. Army Aberdeen Test Center
AVR = automatic voltage regulator
DAS = Data Analysis
EM = electromagnetic
EMI = electromagnetic interference
EMIS = Electromagnetic Induction Spectroscopy
ERDC = U.S. Army Corps of Engineers Engineering Research and Development Center
ESTCP = Environmental Security Technology Certification Program
EQT = Army Environmental Quality Technology Program
GPS = Global Positioning System
HEAT = high-explosive, antitank
JPG = Jefferson Proving Ground
MTADS= Multi-sensor Towed Array Detection Systems
NRL = Naval Research Laboratories
PDOP = precision dilution of precision
POC = point of contact
QA = quality assurance
QC = quality control
ROC = receiver-operating characteristic
RTK = real time kinematic
RTS = Robotic Total Station
SERDP = Strategic Environmental Research and Development Program
UXO = unexploded ordnance
YPG = U.S. Army Yuma Proving Ground

APPENDIX G. DISTRIBUTION LIST

DTC Project No. 8-CO-160-UXO-021

Addressee	No. of Copies
Commander U.S. Army Environmental Center ATTN: SFIM-AEC-ATT (Mr. George Robitaille) Aberdeen Proving Ground, MD 21010-5401	2
Naval Research Laboratory ATTN: (Dr. Herb Nelson) Code 6110 Washington, DC 20375-5342	1
SERDP/ESTCP ATTN: (Ms. Anne Andrews) 901 North Stuart Street, Suite 303 Arlington, VA 22203	1
Commander U.S. Army Aberdeen Test Center ATTN: CSTE-DTC-AT-SL-E (Mr. Larry Overbay)	1
(Library)	1
CSTE-DTC-AT-CS-R Aberdeen Proving Ground, MD 21005-5059	1
Defense Technical Information Center 8725 John J. Kingman Road, STE 0944 Fort Belvoir, VA 22060-6218	2

Secondary distribution is controlled by Commander, U.S. Army Environmental Center, ATTN: SFIM-AEC-ATT.